

Infra- and Transspecific Clues to Understanding the Dynamics of Transposable Elements

Cristina Vieira (✉) · Marie Fablet · Emmanuelle Lerat

Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon; Université Lyon 1;
CNRS; UMR 5558, 69622 Villeurbanne, France
vieira@biomserv.univ-lyon1.fr

Abstract All genomes contain, to a greater or lesser extent, sequences that do not seem to be beneficial. The most preeminent group consists of transposable elements (TEs). These repeated DNA sequences have a significant influence on genome dynamics and evolution. One of the main challenges facing modern molecular evolution is to understand and measure their impact on evolution. The aim of this paper is to establish the relevance and contribution of population studies, as well as the species comparative approaches, to understanding the dynamics of TEs. Most of the examples cited concern the species *Drosophila melanogaster*, since this is one of the genetic key-model organisms, for which an enormous amount of data has been collected over a period of 100 years of genetic research, and which represents a genus for which the genomes of 12 species have been sequenced.

Abbreviations

TE	Transposable element
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
UTR	Untranslated region
RNAi	RNA interference
siRNA	Small interfering RNA
rasiRNA	Repeat-associated small interfering RNA

1

Introduction

Historically, the conventional view of genome evolution has associated organism complexity with the number of protein coding genes. However, the recent complete sequencing of the human genome has shown how similar it is to that of *Drosophila*, since the human genome only has twice its gene number (Lander et al. 2001), and this has highlighted the relevance of non-protein coding gene pathways in controlling the differentiation and diversity of organisms (Taft and Mattick 2003). Another important finding arising from this sequencing program is that only 2% (Goodstadt and Ponting 2006; Human

Genome Sequencing Consortium 2004) of the human genome actually codes for proteins, the function of the rest being still unknown. Large-scale studies have shown that some non-coding regions are very well conserved across species, which suggests that they must in fact have some “function” (Bejerano et al. 2004). In the human genome, 42% of these “non protein-coding gene” regions are constituted by transposable elements (TEs) (Human Genome Sequencing Consortium 2004). This inevitably raises the question of the role of TEs in genome evolution and makes them indwelling components of genomes (see Walisko et al., in this volume).

For a long time scientists thought that the genome was a stable entity, and it was only in the 1950s, thanks to the work of McClintock, that doubts began to trouble this supposedly calm landscape (McClintock 1984, Fedoroff and Botstein 1992). The genome fluidity we now consider as obvious, was difficult to accept at the time. The first TEs were discovered in maize (McClintock 1950), but most of the subsequent early work was done on bacteria (Shapiro 1969, Saedler and Starlinger 1992), since they were a lot easier to study at the molecular level. The first TEs to be described were DNA transposons, i.e., elements that transpose via a DNA intermediate. Studies in *Drosophila*, *Caenorhabditis elegans*, and other eukaryotes, subsequently identified RNA elements, i.e., elements that transpose using an RNA intermediate. Herein, we make no claim to discuss the precise classification of TEs, both because several different systems are possible, and also because new elements are reported every day (Kapitonov and Jurka 2008; Wicker et al. 2007). We will therefore adopt the former classification proposed by Finnegan (1989), which distinguishes two major classes of TEs, based on their transposition cycle intermediates.

TEs are DNA sequences that encode the enzymes necessary for their transposition, i.e., to allow them to move between non-homologous regions in the genomes or to copy themselves to other positions. In some cases, TEs known as non-autonomous sequences do not produce their own enzymes, but are able to use those from functional copies or even from other TE families. The amount of TEs and its impacts on genome stability vary widely among organisms. For instance, retrotransposons constitute almost one half of the human genome, but they are responsible for only 0.2% of spontaneous mutations (Kazazian 1998), while in *Drosophila*, for which the TE contribution is much reduced in terms of genome occupancy, TEs are proposed to be the source of more than 50% of spontaneous mutations with notable effects (Eickbush and Furano 2002). Transposition rates may thus be higher than spontaneous mutation rates, as in *Drosophila*, in which these rates are estimated to be 10^{-3} – 10^{-4} (Vieira and Biémont 1997; Suh et al. 1995; Nuzhdin and Mackay 1995) and 10^{-8} (Crow and Simmons 1983), respectively.¹ The evolution of new

¹ These are global values for the transposition rates, independently of mutation causes such as double-stranded breaks, as suggested by W.D. Heyer in the third volume of this collection.

insertions in a genome should be considered at two time scales. The short-term effects will depend on the insertion site; if the insertion disrupts a gene and consequently affects the fitness of the organism, we can expect it to be eliminated by natural selection, whereas if the insertion is in a non-coding region, we may expect it to be maintained if it has no impact on host fitness.² Long-term effects will only involve insertions that are associated with very weak deleterious effects (Langley et al. 1988), since these are the only ones not promptly eliminated. This makes it possible to identify fixed insertions in populations, which may not necessarily be adaptive, but can simply be the consequence of genetic drift and bottlenecks (Cordaux et al. 2006a; Cordaux et al. 2006b). Furthermore, insertions of TEs may modify regulatory pathways and the expression patterns of genes when they insert in their vicinity (Peaston et al. 2004), and may also be subject to strong selection, leading to an increase in their frequency in populations and enhanced host fitness (Aminetzach et al. 2005). The occurrence of molecular domestication events is now frequently reported, and seems to happen in many different organisms (Feschotte and Pritham 2007; Kapitonov and Jurka 2005; Miller et al. 1997), implying that TEs play a key role in genome evolution.³

We describe here the way population-based studies and species comparative analyses have contributed to the current understanding of TE dynamics and evolution, focusing on different levels of study of TEs, from the copy number, to sequence variation, and the epigenetic regulation of activity.

2

Lessons from the Past

2.1

The Heritage of Hybrid Dysgenesis Studies in *Drosophila* Populations

After their discovery by Barbara McClintock in the 1950s, TE study went through a new birth in the late 1970s when drosophilists related aberrant traits in some crosses of *Drosophila melanogaster* strains. Among these aberrant traits were recombination in males (Hiraizumi et al. 1973) – which is not expected to occur in *D. melanogaster* –, high rate of mutation (Thompson and Woodruff 1980), sterility, chromosomal aberrations (Kidwell et al.

² Here, we only refer to “regular”, punctual transposition events, as opposed to the massive bursts of transposition observed in *Drosophila* in the case of what is called hybrid dysgenesis (Kidwell et al. 1977), which will be developed in Sect. 2.1. This phenomenon is observed when crossing individuals originating from strains differing in their TE content, and results in a high rate of mutation, chromosome rearrangements, and sterility in the offspring, due to an extremely elevated rate of transposition. In this case, even if the insertion sites are not located in coding regions, the effects on the offspring fitness are considerable.

³ For an extensive review on domestication of TEs, refer to Dettai and Volff, in this volume, and Volff 2006.

1977). These aberrations were found non reciprocally in F1 hybrids, and some of the traits were even not found in non hybrids. This led Margaret Kidwell and colleagues (1977) to use the term “hybrid dysgenesis” to qualify such a phenomenon. *D. melanogaster* strains could be classified into two types, called P and M, according to the paternal or maternal contribution in the production of hybrid dysgenesis. It appeared that strains collected from natural populations at that time were typically of the P type and those having a long laboratory history were of the M type (Kidwell et al. 1977). At the same time, Picard (1976) reported another system of hybrid dysgenesis, distinguishing inducer (I), reactive (R), or neutral (N) strains. All strains collected from the wild were classified as I strains. Geneticists at that time proposed that all of these aberrant traits could be related, and caused by chromosomal factors, but their identification proved to be hard due to the difficulty in localizing the causal factor(s) to a single chromosome (Kidwell et al. 1977). It was subsequently considered that hybrid dysgenesis in the P-M system resulted from the interaction of a chromosomal component (“P factor”) and an extrachromosomal property (“M cytotype”) (Engels and Preston 1980). This P factor actually corresponds to the now well-studied P transposon, and the I-R hybrid dysgenesis proved to be due to another transposable element, the I non-LTR retrotransposon.

Studies of the hybrid dysgenesis phenomenon proved that the invasion of a genome by TEs was possible and could happen in a relatively short time. This motivated the approach of TEs by modeling, so that in the early 1980s, several authors proposed theoretical models intended to explain the dynamics of TEs (Le Rouzic and Decelière 2005 for a review). These relatively simple models could be used to test neutrality or selection of the deleterious effects of TE insertion, or the effects of recombination induced by TEs. The value of a model depends on being able to test it. In this respect, *Drosophila* is a very suitable model organism for such tests. In fact, *Drosophila* is unmatched in two characteristics: (1) the giant polytene chromosomes and (2) balancer chromosomes. One most prominent experimental tool distinguishing *Drosophila* from other model systems is the advantage to be able to carry out in situ hybridizations on polytene chromosomes (Gall and Pardue 1969; Pardue and Gall 1969), and map the sites as well as determine the copy number (Biémont et al. 2004, Fig. 1) by means of the classical and still very useful chromosome maps of Bridges (Bridges 1935).

The first studies were done on laboratory populations of *Drosophila* (Langley et al. 1988), and then several studies were performed on natural populations (Biémont 1994; Biémont et al. 1994; Hoogland and Biémont 1996). As has been demonstrated in several reviews, no general model can be applied to all TEs and all populations, since they both are rarely at equilibrium (Biémont et al. 1997). Further, the precise biochemical details of transposition of a TE family in general and each individual TE specifically, embedded in its particular chromatin environment, is different in each specific circumstance. Copy number

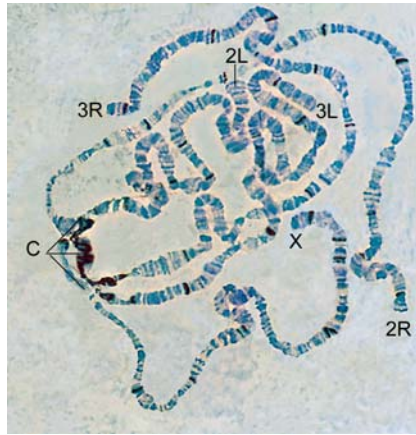


Fig. 1 In situ hybridization on *Drosophila* salivary gland polytene chromosomes. The hybridization was performed with a biotinylated DNA probe (reviewed by Biémont et al. 2004). The probe, with sequence homology to the 412 LTR retrotransposon, is detected as multiple black bands on the chromosome preparation. The position of each TE can be precisely identified and linked to the maps of the complete *Drosophila* genome. C: chromocenter, 2L and 2R are the left and right arms of the chromosome 2, 3L and 3R left and right arms of the chromosome 3

data obtained by in situ hybridization in *D. melanogaster* were not easy to extrapolate to other species of *Drosophila*, even to closely related species such as *D. simulans* (Vieira and Biémont 1996, 2004; Vieira et al. 2000). Comprehensive analyses of numerous individuals and populations soon became impossible to manage practically. In addition, one of the main problems with the in situ approach is the approximate nature of the localizations. It is quite difficult to distinguish between neighboring sites, and also to be sure of the sequence similarity between the probes and the highlighted spots. This made it impossible to identify all potentially fixed sites, leading to the conclusion that insertion polymorphism levels in *Drosophila* were high. Using the insertion sites detected in the sequenced genome and searching for them in individuals in a natural population, led to the identification of numerous fixed insertions. The evolutionary significance of these insertions is still under investigation, and we need to be able to distinguish between genetic drift and adaptive selection (Aminetzach et al. 2005; Lipatov et al. 2005; Macpherson et al. 2008; McCollum et al. 2002; Dettai and Volff, in this volume).

2.2

The Sibling Species *D. melanogaster* and *D. simulans*

As previously mentioned herein, the number of TE copies varies extensively when considering different model genomes, such as *D. melanogaster*, *Homo*

sapiens, or *Zea mais*. Nevertheless, one might have expected that closely related species had the same copy number of TEs – at least of the same order of magnitude, if we assume that these species have been submitted to similar evolutionary processes. However, this assumption turned out to be incorrect, as was revealed by the analysis of two sibling species of the genus *Drosophila*, *D. melanogaster* (the model species of metazoan genetics) and *D. simulans*. It has been shown that the copy number of most TEs (obtained by in situ hybridization) is smaller in *D. simulans* than in *D. melanogaster*. But more surprisingly, there is a huge difference in copy numbers between natural populations of *D. simulans* with regard to several TE families. In fact, most populations have very low copy numbers, but there are a few exceptions, in which the copy number is very high, sometimes even higher than the average value found in *D. melanogaster*. This has led us to hypothesize that the genome of *D. simulans* is beginning to be invaded by TEs, and that this invasion could be associated with the current worldwide colonization of the *D. simulans* species (Biémont et al. 2003; Vieira and Biémont 2004; Vieira et al. 1999). However, the alternative hypothesis of an ancient invasion followed by a progressive loss of TEs cannot be ruled out. Recent data obtained for a LINE element and an LTR retrotransposon seem to support the latter hypothesis (Fablet et al. 2006; Rebollo et al. 2008). The main challenges facing us are understanding why some populations are sporadically invaded by a specific family of TEs, identifying the genetic and/or environmental factors that have allowed this invasion, and finding out how TEs are eliminated.

2.3

In the Genome Sequencing Era

The recent explosion of sequencing projects, making ever more genomes available, is an important step forward in determining the TE loads of different species. The analysis of the genomes of 12 *Drosophila* species and the genomes from other insects, such as *Anopheles gambiae* (Holt et al. 2002), *Aedes aegypti* (Nene et al. 2007), *Pediculus humanus*, *Bombyx mori* (Xia et al. 2004), *Tribolium castaneum* (Tribolium Genome Sequencing Consortium 2008), *Nasonia vitripennis*, or *Apis mellifera* (Honeybee Genome Sequencing Consortium 2006), has shown that there are significant differences in the amount, type and degree of conservation of TEs between different species. The genome of *Apis* differs from previously sequenced insect genomes in that it presents very small amounts of TEs, with especially very few retrotransposons (Honeybee Genome Sequencing Consortium 2006). Most of the TEs in *Apis* are from the *mariner* family, a DNA transposon, whereas other types of transposons and retrotransposons are present, but only as highly degraded copies, indicating that they are no longer active. In contrast, the silkworm has a very large genome, of which TEs account for 21.1%, and it is probably TEs that are responsible for the increase in genome size in this species (Xia

et al. 2004). Most of these TEs mainly belong to one family of LTR retrotransposons. The genome of *Anopheles* is also larger than that of *D. melanogaster*, and in this species TEs account for about 16% of the euchromatin with a good representation of different families from DNA transposons and retrotransposons (Holt et al. 2002).

The analyses of other eukaryotic genomes have demonstrated very different patterns of TE dynamics. For example, traces of waves of amplification/loss over evolutionary time have been identified in the human genome, where the *L1* elements (non-LTR retrotransposons) are the ones that have invaded the genome the most recently. Furthermore, several observations (Boissinot and Furano 2005; Mathews et al. 2003) indicate that competition exists between different LINE 1 families, which could explain why only one subfamily is now active in the human genome. The amplification of TEs has been reported in many plants, and is responsible for the increase in the genome size of maize (SanMiguel et al. 1998).

However, despite the large amount of data that can be obtained from analyses of sequenced genomes, specific population studies addressing specifically particular TEs and their molecular characteristics of transposition are still the only way to approach some particular issues. For instance, Herrera et al. (2006) recently used PCR analysis to investigate the presence/absence of four insertions of the *HERV-K* element in nine human populations from various geographical origins. The authors observed a general discrepancy from Hardy–Weinberg equilibrium expectations, which they propose might be due to the fact that the heterozygotes are subjected to pandemic negative selection. These observations are of great interest in attempting to understand the selective pressures that drive TE dynamics within a germline genome.

3

Towards an Understanding of TE Regulation. From Sequence to Epigenetics

3.1

Sequence Variability

The other way to analyze TE dynamics is to compare their intrinsic sequence characteristics. This allows us to address questions regarding the phylogeny and evolution of individual TEs. These types of analyses are based on the assumption that the evolution rates of TEs are the same as those of conventional genes, which of course is not always true (Malik et al. 1999; Malik and Eickbush 1999, 2001). One of the most striking conclusions of the phylogeny studies from sequenced genomes was that TEs evolve independently of their host genomes (Lerat et al. 2002a,b). Phylogenetic studies of the same TEs in different species have demonstrated that horizontal trans-

fers can occur (Clark et al. 1994). These studies were mainly based on coding regions, and there have been very few population studies (Silva et al. 2004).

On the other hand, several population studies have been carried out on non-translated regions, such as the long terminal repeat (LTR) and untranslated region (UTR). LTR-retrotransposons are the most common TEs in eukaryotic organisms, especially in *Drosophila*, in which more than 2/3 of the TEs belong to this class (Hoskins et al. 2007). These TEs transpose via an RNA intermediate, and are dependent on reverse transcriptase activity. The nature of the replication process of these retrovirus-like TEs favors the appearance of duplications that may behave as enhancers during the evolutionary process (McDonald et al. 1997). Such enhancers may act in *cis* on the element itself, or may influence other genes in the vicinity of their point of insertion (Peaston et al. 2004). In recent years several studies have attempted to characterize the regulatory regions of some elements. In *Drosophila*, TEs such as *copia*, *blood*, 412, *gypsy*, *tirant*, *micropia*, *Ulysses*, and *ZAM* have been extensively studied, and in some cases natural population lines were used. For example, the *copia* element has repetitive motifs in its 5' LTR and 5' UTR⁴ regions in *D. melanogaster* (Matyunina et al. 1996), and this region acts as an enhancer in regulating the expression of the element (Wilson et al. 1998). An analysis of several natural populations of *D. melanogaster* and *D. simulans* has revealed the existence of three types of sequence variants in the region spanning from the 5' LTR to the 5' UTR, depending on whether specific portions of the sequence that encompass repeated motifs have been deleted. Two of these *copia* variants are present in both species, but the third occurs only in *D. simulans*. The existence of regulatory region variants has also been reported later for other TEs, such as *blood* (Costas et al. 2001), 412 (Mugnier et al. 2005), and *tirant* (Fablet et al. 2006). In some cases, all the variants seem to be active, but the activity of others such as *copia* and *tirant*, seems to be restricted to a few populations in *D. simulans*.

D. melanogaster populations tend to have rather comparable TE contents (Vieira et al. 1999), but they display considerable variability in their TE regulatory region sequences. For instance, the 5' UTR region of the *tirant* element displays a high variability with regard to the number of tandem repeats of a 102-bp motif among *D. melanogaster* populations worldwide. In addition, investigation of the sequence variability of the *tirant* regulatory region in the four closely related species of the *D. melanogaster* subgroup (*D. melanogaster*, *D. simulans*, *D. sechellia*, and *D. mauritiana*) has revealed different pathways for *tirant* dynamics and evolution, mainly depending on the structure of the populations and on the history of the species (Fablet et al. 2007b). A classi-

⁴The typical structure of LTR retrotransposons is made of the genes necessary for their mobilization, bordered by two LTRs, 5' and 3'. Just downstream of the 5' LTR, and 3' to the tRNA primer binding site, a transcribed-untranslated region (UTR) is found. Regulatory sequences are concentrated in LTR and this UTR region (Fablet et al. 2007a).

cal example is the *micropia* retrotransposons of *D. melanogaster* and *D. hydei* (Lankenau et al. 1988; Huijser et al. 1988). While the LTRs of *micropia* within the same species are highly conserved between individual *micropia* elements, the LTRs of *micropia* elements from the other species are completely different except a tandem of putative ecdysteroid receptor binding sites (Lankenau et al. 1990).

Once genomes began to be sequenced, global studies of the amount and structure of the TEs and their dynamics within a given family were possible. The first eukaryotic genome to be sequenced (in 1996) was the yeast *Saccharomyces cerevisiae*, and this was the beginning of global genomic analyses of TEs (Bassett et al. 1996). In particular, Jordan and McDonald (Jordan and McDonald 1998, 1999) analyzed the different families of the *Ty* LTR retrotransposons, and showed that they had recently transposed, and were subject to severe selective pressure. From the analysis of the sequenced genome of *D. melanogaster* (Bowen and McDonald 2001; Lerat et al. 2003), we have learned that the euchromatic insertions of TEs are very recent in this species, and that there is no continuous range of divergence among copies in each class. Indeed, most of the TEs are very similar in sequence to the canonical ones, whereas the other few TEs display high divergence. On the basis of this observation, the hypothesis currently accepted is that in *D. melanogaster*, after intensive TE activity, any degenerate elements are eliminated by a high turnover mechanism. There are of course exceptions, since some very old elements can also be found such as *helena* (Rebollo et al. 2008) or *micropia* (Lankenau et al. 1988). The question as to whether this high turnover mechanism is shared by other *Drosophila* species should soon be resolved, since 11 other *Drosophila* genomes have been recently sequenced (Drosophila 12 Genomes Consortium 2007). Indeed, the analysis of the *D. simulans* sequenced genome has shown, surprisingly, that contrarily to the in situ hybridization data obtained before, this species contains a high number of deleted and rearranged TEs, with few complete sequences for a few families of TEs, which were probably horizontally transmitted into the *D. melanogaster* genome. Moreover, no high turnover is detected in *D. simulans* (Lerat et al., unpublished data).

The initial sequencing programs concentrated on the euchromatic portion of the genomes, since this is the region where active genes have usually been identified. Furthermore, the small amount of highly repeated sequences in these regions makes them much easier to sequence and annotate. Recently, *Drosophila* heterochromatin was sequenced, and it was found that 77% of it is composed of TEs, with few complete elements (Hoskins et al. 2007). Analysis of the LTR proviral-like elements revealed the presence of numerous full-length sequences, which indicates invasion of heterochromatin to be a recent phenomenon (Mugnier et al. 2008). The dynamics of heterochromatin expression is still unknown, but we can envisage several types of gene regulation that could take into account the

high proportion of repeated sequences. In addition, it has also been recently demonstrated that some retrotransposons harbored chromodomains in their integrase sequences, which triggers them to heterochromatic regions. These copies then themselves become heterochromatic targets for new subsequent retrotransposon insertions, in an autocatalytic-like fashion (Gao et al. 2008).

The significant efforts devoted to genome sequencing have increased the relevance of comparative genomic approaches, and made it easier to perform gene prediction and annotation. However, from a population genomics point of view, the sequencing of different species is definitely not enough. The case of *D. melanogaster* is particularly striking, as the sequenced strain, an M strain already used by Morgan in the first genetic studies on this species (Morgan et al. 1915), corresponds to an ancient laboratory line, extinct in the wild. Its genome is clearly different from those found in individuals from natural populations. For example, this genome does not contain any P element insertion, as the strain was collected before the invasion of the genome by this element, whereas all recent natural populations do contain this TE (Engels 1997). We would expect that, in the near future, sequencing projects will consider several individuals of the same species taken from different populations. This approach has already begun with the sequencing of seven lines of *D. simulans* (<http://www.dpgp.org/>). The first analyses have made it possible to study polymorphism and divergence in this species using a whole genome approach (Begun et al. 2007).

3.2

TE Dynamics at the Epigenetic Level

Epigenetics is an “old” field that is currently experiencing a new flowering.⁵ The number of phenomena that may be considered as epigenetic is constantly on the increase. We now realize that all the different phenomena independently reported some years ago, such as chromatin compaction via DNA methylation, or RNA interference, are actually closely linked epigenetic mechanisms. Considerable evidence has been amassed showing that the regulation of TEs (see Fig. 2) is mediated by epigenetic factors, and that this regulation can impact on the regulation of the host gene (Brennecke et al. 2007; Conley et al. 2008; Weil and Martienssen 2008 for a review). Several examples have also shown how TE regulation can depend on environmental conditions. For example, the transposition of P elements in *Drosophila* is temperature sensitive, and indirectly, copy number variation may be associated

⁵ For an historical perspective on the evolution of the concept of epigenetics, see Jablonka and Lamb 2002. The term is attributed to Conrad Waddington, who used it in the 1940s to refer to “the causal interactions between genes and their products which bring the phenotype into being”. The term was then little used in the following decades and reappeared recently, with a slightly different meaning. It now refers to reversible, heritable modifications of genetic expression that are not due to changes in the DNA sequence.

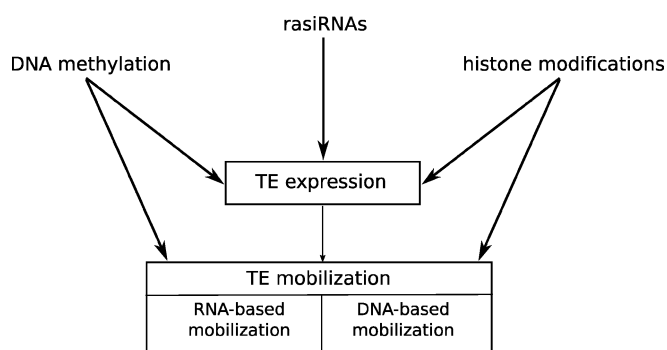


Fig. 2 Schematic view of the different aspects of epigenetic regulation of TEs. Although mechanisms and effects on genome dynamics differ for RNA- and DNA-based TEs, in both cases, mobilization is dependent on TE expression. Several epigenetic pathways may interfere with this expression, such as methylation of cytosines in the DNA molecule, post-translational changes in the histone tails, and RNA interference. These different pathways are interdependent, and a high number of interactions has been described. These epigenetic modifiers have direct effects on gene expression and are responsible for phenotypic changes. In most of the cases, epigenetic modifications are sensitive to the environment, so that genotype X environment interactions can be observed

with environmental factors (Kalendar et al. 2000; Kidwell 1977; Vieira et al. 1998). We are thus tempted to envisage links between epigenetic regulation, TEs and environment. The investigation of natural populations may provide a useful way to address these questions. We will describe here very briefly the three major epigenetic mechanisms known to be implicated in the epigenetic control of TE dynamics.⁶

3.2.1

DNA Methylation

DNA methylation is considered to be the first level of gene regulation. It is an epigenetic mark that is stable and lasting, and it is transmitted throughout development (Reik 2007). In human, *de novo* methylation is mediated by Dnmt3 (Dnmt3a and Dnmt3b) and methylation is then maintained by Dnmt1. A third family has been described, Dnmt2, which methylates asymmetric cytosines and also presents tRNA methyltransferase activity (Jeltsch et al. 2006). The consequences of cytosine methylation depend on the organism concerned and different types of enzymes are involved depending on the species (see Table 1). For example, in mammals, DNA methylation drives gene imprinting during development, leading to the differential expression of the genes (Hellman and Chess 2007). In plants, DNA methylation

⁶ An extensive network of researchers working on the epigenome has been established (the Epigenome Network of Excellence). See <http://www.epigenome-noe.net/> for abundant data on the epigenetic control.

Table 1 Main enzymes involved in DNA methylation in different groups of organisms

	Enzyme	5meCy rate	Methylation type	Reference
Mammals	DNA methyltransferase family (Dnmt1, 2 and 3)	2–5%	Symmetric	Pradhan and Esteve 2003
Plants	Domains rearranged methyltransferase 2 (Drm2) Methyltransferase 1 (Met1) Dnmt2 Chromomethylase (Cmt3)	19%	Symmetric and asymmetric	Chan, Henderson and Jacobsen 2005
Honeybee	Dnmt family	< 1%	Symmetric	Wang et al. 2006
<i>Drosophila</i>	Dnmt2	< 1%	Asymmetric	Marhold et al. 2004
Zebrafish	Dnmt family	1%	Symmetric	Mhanni and McGowan 2004
Nematode	Not identified	ND		Ponger and Li 2005
Yeast	Not identified	ND		Ponger and Li 2005

is mainly associated with TE silencing (Slotkin and Martienssen 2007). Until recently, *Drosophila* was considered to be a non-methylated organism, in contrast to most vertebrates and insects. Actually, *Drosophila* does methylate, but methylation is restricted to young embryos (Lyko 2001; Lyko et al. 2000), and mainly to cytosines associated with A and T nucleotides, which contrasts with the situation in the other organisms, in which methylation takes place on CpG dinucleotides. In addition, unlike other organisms, the methyltransferase responsible for methylation in *Drosophila* is Dnmt2, and the overall level of genome methylation is low. The impact of methylation in the *Drosophila* genome is still not well understood, but we can speculate that even if the levels of methylation are low, they could be sufficient to be involved in the regulation of TEs. Other insects such as the haplo/diploid honey bees (*Apis mellifera*) encompass a complete CpG methylation program (similar to humans) and might be a useful system to study epigenetic control of TEs in the future (Schaefer and Lyko 2007; Kucharski et al. 2008).

Global analysis of genome methylation rates are necessary to understand the real evolutionary impact of DNA methylation. High-performance capillary electrophoresis (HPCE) or liquid chromatography (LC) are common techniques to determine these global rates (Berdasco et al. 2009). Sequence information can be obtained by bisulfite techniques, which can be associated with methylation-sensitive restriction enzymes (Suzuki and Bird 2008, for a review). We can expect that in a near future methylomes from different species, but also from different individuals, will be available.

3.2.2

Histone Modifications

In the cell nucleus, DNA is associated with proteins (such as histones) to form the chromatin. Chemical modifications of the N-terminal as well as C-terminal histone tails will make the DNA more or less accessible to the transcription machinery. Thus, the post-translation state of the histones will regulate both gene activity (Buratowski and Moazed 2005) and TE activity (Hall et al. 2002; Tran et al. 2005). Histone modifications are considered to be reversible, sensitive cellular tools of gene regulation, able to control genes during development, turning them ON or OFF. The histone pattern of modifications thus represents the expression of genes and TEs in a genome (Gendrel et al. 2002; Goldberg et al. 2007; Martens et al. 2005), even if in *Drosophila*, the relationship with TE expression has still not been clearly demonstrated. If a TE is inserted in an open conformation of chromatin, invasion of the genome by this TE is possible. On the contrary, if the TE is inserted in closed chromatin, the transposition will be difficult and genome invasion less likely to occur.

3.2.3

RNA Interference (RNAi)

The existence of antisense RNAs produced by TEs have previously been reported and proposed to play a role in the control of their activity (for instance Lankenau et al. 1994 for the *micropia* LTR retrotransposon; Simmons et al. 1996 for the P transposon). Variability among strains in the efficiency of such a regulation was already proposed, underlying the need for a populational approach of these issues (Simmons et al. 1996).

More recently, small antisense RNA molecules have been demonstrated to play multiple roles in the regulation of gene expression (Fire et al. 1998). These include (1) the degradation of mRNA by small interfering RNA (siRNA), which leads to post-transcriptional repression of the genes (Post-Transcriptional Gene Silencing: PTGS), (2) repression of the translation of mRNA by micro RNA (miRNA) and (3) the repression of the transcription (Transcriptional Gene Silencing: TGS) (Buchon and Vaury 2006). This is a complex field of research, and various different pathways have been identified that lead to RNAi (Matranga and Zamore 2007). In *Drosophila*, a new type of siRNAs has just been described: “repeated-associated small interfering RNAs” or rasiRNAs. These entities are homologous to the antisense RNAs of some TE families (Aravin et al. 2007; Kavi et al. 2005; Pélisson et al. 2007; Vagin et al. 2004, 2006; Vaughn et al. 2007). This pathway involves Piwi-like proteins, members of the Argonaute family, not only in the germline but also in somatic cells (Kawamura et al. 2008).

The discovery of these rasiRNAs shed some significant light on the already identified control loci *flamenco* or COM, but whose precise molecular role was so far not understood. The heterochromatic regulatory locus COM, involved in the control of the *ZAM* and *Idefix* LTR retrotransposons, has indeed been demonstrated to direct TE somatic silencing through Piwi-dependent and -independent pathways (Desset et al. 2008). COM is located in the same region as the *flamenco* locus, which controls the *gypsy* LTR retrotransposon, also via the production of rasiRNAs (Pélisson et al. 2007). Brennecke et al. (2007) recently demonstrated that *flamenco* was a piRNA cluster, from which TE homologous piRNAs were produced. More recently, the P-M system and the I-R system have been shown to be entirely linked to the RNAi regulation (Brennecke et al. 2008).

In this regard, attention is paid to an intriguing electron-dense structure surrounding the nucleus of germline cells in *Drosophila*, called “nuage” (French word for “cloud”). Its exact function is still unknown, but it is now known that the molecular actors of the production of rasiRNAs localize into the nuage, which is therefore assumed to be a crucial structure in the regulation of TE expression (Lim and Kai 2007; Pane et al. 2007; see also Lankenau, vol. 3, this series). The germline-specific nuage of *Drosophila* seems to be conserved in other organisms. For instance, it corresponds to the chromatin bodies of mammals, which also appears to be involved in the germline epigenetic control of TEs (Soper et al. 2008).

RNAi is implicated in a huge number of phenomena, and this must be integrated in the analysis of natural populations.

3.2.4

Population Epigenomics

It is becoming obvious that some epigenetic modifications are heritable (Cavalli and Paro 1999; Goldberg et al. 2007), although the mechanism by which this transmission is achieved remains unknown (Cuzin et al. 2008). The main challenge for geneticists is to understand the impact of this epigenetic code. Do different populations have different epigenetic pathways? What are the consequences of epigenetics for evolution? How does the genome take the changes in epigenetic states into account? Can we attribute differences between populations to differences in their epigenetic status? And finally, could differences in TE activity (transposition) be based on differing epigenetic controls? Very few studies have been done to analyze the epigenetic variability of TE regulation in natural populations, but some pioneer work has been performed in plants. In *Arabidopsis*, Vaughn and colleagues have recently shown that the cytosine methylation rate on genes differed depending on the ecotype analyzed, whereas no corresponding difference was detected in TE methylation (Vaughn et al. 2007). More recently, Zhai et al. (2008) have shown that these two ecotypes have different levels of siRNA production, implying

that the siRNA cluster may not be the same in different populations. Similar studies have not yet been carried out in other organisms, although the data obtained from comparisons of the epigenome of monozygotic twins have shown how important the epigenome may be in determining the expression of the phenotype (Fraga et al. 2005).

Drosophila represents an excellent model for the study of population epigenomic variability, because of its relatively low – compared to some mammals or plants – number of TEs, the large number of mutant strains, and the availability of natural populations. Global methods for the analysis of the chromatin structure (Chip seq), for the analysis of the methylome and for the sequencing of small RNAs are now popular. Comparison of populations living in different environments and having different TE features (Vieira et al. 1999) is now easier, and does not even imply the artificial construction of the experiments. Population analysis of the epigenomes may be one of the ways used in the future to understand how individuals adapt to new environments, and shed new light on the study of genotype X environment interactions. In particular, some works have demonstrated the heritability of epigenetic marks (Cavalli and Paro 1999; Richards 2008).

4

Conclusion

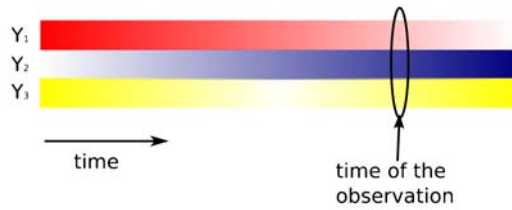
Whole genome sequencing and population analyses should be viewed as complementary methods, each focusing on a different aspect of TE dynamics (see Fig. 3). Whole genome sequence analysis can be used to compare different families of TEs within a given genome, while population sequence analysis compares different genomes with regard to a given family of TEs.

Evolution is (predominantly) a stochastic process. The scale of impact of stochasticity depends on the constraints applied to a given section of the genome: the greater the constraints, the narrower the range of possible observable outcomes of evolution. This is one of the reasons why the variability observed for the TE load of a genome is much higher than that observed for the pool of protein coding genes (because weaker constraints act on TEs). This means that population-based studies can take into account the variability of pathways authorized by this combination of stochasticity and weaker constraints, whereas a single sequenced genome focuses on only one of the multitude of possible states.

The final alliance between sequenced genomes and population studies might be provided by the new sequencing techniques, such as 454 (Margulies et al. 2005), which will provide rapid sequencing of full genomes from different populations, and even from different individuals of a given species. The 454 technology, which is already very frequently used for the study of “non-model” organisms, might at first not appear to be ideal for the study

Y_1, Y_2, Y_3 :
studied variables,
such as amount of TE,
copy number,
methylation level, etc.

A. The sequenced genome



B. Populations

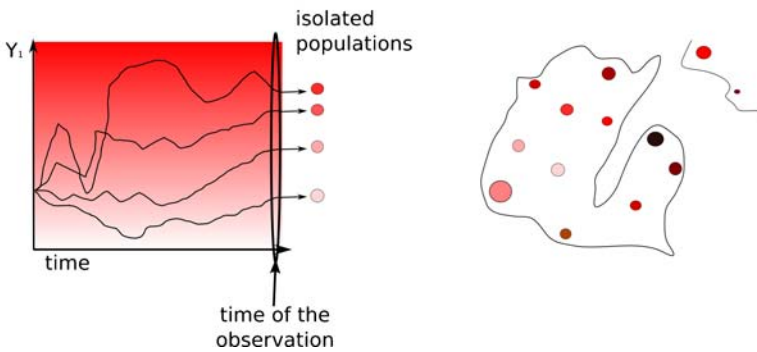


Fig. 3 Different, yet complementary, points of view provided by the study of the sequenced genome of a species (**A**), and several geographically distinct populations (**B**). The figure is drawn for any variable of study, called Y for convenience, such as for instance TE amount in the genome, copy number for a given TE family, level of DNA methylation in TE sequences, etc. Color variation from light to dark: variation of Y value from minimum to maximum. **A**, whole genome sequencing of an individual organism. The amount of data extracted is extensive since several Y variables (Y_1, Y_2, Y_3) can be studied at the same time. However, it provides only one value for each Y considered, without any information on their possible ranges of variation, contrary to populational studies. **B Left panel**: temporal evolution of the populations regarding to the value for Y_1 . Several populations split from the original population, and start to evolve separately. With time, the values of Y_1 evolve differently in each population, so that at a given time of sampling, the populations are different regarding Y_1 . **B Right panel**, spatial representation of populations and their corresponding values for Y_1 . Isolated populations give rise to individual genomic samples. Each population displays a different value for Y_1 .

of repeated DNA, since the fragments sequenced are very short, but some studies have already been done and look promising (Macas et al. 2007).

Acknowledgements We are grateful to Christian Biémont for critical reading of the manuscript and to Monika Gosh for English correction. This work was supported by CNRS and Finovi.

References

- Aminetzach YT, Macpherson JM, Petrov DA (2005) Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* 309:764–767
- Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318:761–764
- Bassett DE Jr, Basrai MA, Connelly C, Hyland KM, Kitagawa K, Mayer ML, Morrow DM, Page AM, Resto VA, Skibbens RV, Hieter P (1996) Exploiting the complete yeast genome sequence. *Curr Opin Genet Dev* 6:763–766
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD et al (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5, e310
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D (2004) Ultraconserved elements in the human genome. *Science* 304:1321–1325
- Berdasco M, Fraga ME, Esteller M (2009) In: Tost J (ed) Quantification of global DNA methylation by capillary electrophoresis and mass spectrometry DNA methylation methods and protocols, 2nd edn. Springer Protocols, The Netherlands, pp 23–34
- Biémont C, Monti-Dedieu L, Lemeunier F (2004) Detection of transposable elements in *Drosophila* salivary gland polytene chromosomes by in situ hybridization. *Methods Mol Biol* 260:21–28
- Biémont C, Nardon C, Decelière G, Lepetit D, Loevenbruck C, Vieira C (2003) World-wide distribution of transposable element copy number in natural populations of *Drosophila simulans*. *Evolution* 57:159–167
- Biémont C, Tsitrone A, Vieira C, Hoogland C (1997) Transposable element distribution in *Drosophila*. *Genetics* 147:1997–1999
- Biémont C (1994) Dynamic equilibrium between insertion and excision of P elements in highly inbred lines from an M' strain of *Drosophila melanogaster*. *J Mol Evol* 39:466–472
- Biémont C, Lemeunier F, Garcia Guerreiro MP, Brookfield JF, Gautier C, Aulard S, Pasyukova EG (1994) Population dynamics of the copia, mdg1, mdg3, gypsy, and P transposable elements in natural populations of *Drosophila melanogaster*. *Genet Res* 63:197–212
- Boissinot S, Furano AV (2005) The recent evolution of human L1 retrotransposons. *Cytogenet Genome Res* 110:402–406
- Bowen NJ, McDonald JF (2001) *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* 11:1527–1540
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128:1089–1103
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322:1387–1392
- Bridges CB (1935) Salivary chromosome maps: with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J Hered* 26:60–64
- Buchon N, Vaury C (2006) RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity* 96:195–202

- Buratowski S, Moazed D (2005) Gene regulation: expression and silencing coupled. *Nature* 435:1174–1175
- Cavalli G, Paro R (1999) Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* 286:955–958
- Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* 6:590
- Clark JB, Maddison WP, Kidwell MG (1994) Phylogenetic analysis supports horizontal transfer of P transposable elements. *Mol Biol Evol* 11:40–50
- Conley AB, Miller WJ, Jordan IK (2008) Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet* 24:53–56
- Cordaux R, Hedges DJ, Herke SW, Batzer MA (2006a) Estimating the retrotransposition rate of human Alu elements. *Gene* 373:134–137
- Cordaux R, Lee J, Dinoso L, Batzer MA (2006b) Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene* 373:138–144
- Costas J, Valade E, Naveira H (2001) Amplification and phylogenetic relationships of a sub-family of blood, a retrotransposable element of *Drosophila*. *J Mol Evol* 52:342–350
- Crow JF, Simmons MJ (1983) The mutation load in *Drosophila*. In: Carson HL, Ashburner M, Thomson JN (eds) *The genetics and biology of Drosophila*. Academic Press, London, pp 1–35
- Cuzin F, Grandjean V, Rassoulzadegan M (2008) Inherited variation at the epigenetic level: paramutation from the plant to the mouse. *Curr Opin Genet Dev* 18:193–196
- Desset S, Buchon N, Meignin C, Coiffet M, Vaury C (2008) In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and independent pathways. *PLoS ONE* 2:e1526
- Drosophila* 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218
- Eickbush TH, Furano AV (2002) Fruit flies and humans respond differently to retrotransposons. *Curr Opin Genet Dev* 12:669–674
- Engels WR (1997) Invasions of P elements. *Genetics* 145:11–15
- Engels WR, Preston SR (1980) Components of hybrid dysgenesis in a wild population of *Drosophila melanogaster*. *Genetics* 95:111–128
- Fabiet M, McDonald JF, Biémont C, Vieira C (2006) Ongoing loss of the tirant transposable element in natural populations of *Drosophila simulans*. *Gene* 375:54–62
- Fabiet M, Rebollo R, Biémont C, Vieira C (2007a) The evolution of retrotransposon regulatory regions and its consequences on the *Drosophila melanogaster* and *Homo sapiens* host genomes. *Gene* 390:84–91
- Fabiet M, Souames S, Biémont C, Vieira C (2007b) Evolutionary pathways of the tirant LTR retrotransposon in the *Drosophila melanogaster* subgroup of species. *J Mol Evol* 64:438–447
- Fedoroff N, Botstein D (1992) *The dynamic genome* Barbara McClintock's ideas in the century of genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331–368
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5:103–107
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811
- Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E,

- Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 102:10604–10609
- Gall JG, Pardue ML (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci USA* 63:378–383
- Gao X, Hou Y, Ebina H, Levin HL, Voytas DF (2008) Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res* 18:359–369
- Gendrel AV, Lippman Z, Yordan C, Colot V, Martienssen RA (2002) Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene DDM1. *Science* 297:1871–1873
- Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell* 128:635–638
- Goodstadt L, Ponting CP (2006) Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput Biol* 29:e133
- Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI (2002) Establishment and maintenance of a heterochromatin domain. *Science* 297:2232–7
- Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. *Science* 315:1141–3
- Herrera RJ, Lowery RK, Alfonso A, McDonald JF, Luis JR (2006) Ancient retroviral insertions among human populations. *J Hum Genet* 51:353–362
- Hiraizumi Y, Slatko B, Langley C, Nill A (1973) Recombination in *Drosophila melanogaster* male. *Genetics* 73:439–444
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG et al (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298:129–149
- Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931–49
- Hoogland C, Biémont C (1996) Chromosomal distribution of transposable elements in *Drosophila melanogaster*: test of the ectopic recombination model for maintenance of insertion site number. *Genetics* 144:197–204
- Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F, Villasante A, Dimitri P, Karpen GH, Celniker SE (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* 316:1625–1628
- Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945
- Huijser P, Kirchhoff C, Lankenau DH, Hennig W (1988) Retrotransposon-like sequences are expressed in Y chromosomal lampbrush loops of *Drosophila hydei*. *J Mol Biol* 203:689–697
- Jablonka E, Lamb MJ (2002) The changing concept of epigenetics. *Ann NY Acad Sci* 981:82–96
- Jeltsch A, Nellen W, Lyko F (2006) Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 31:306–308
- Jordan IK, McDonald JF (1999) Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. *Genetics* 151:1341–1351
- Jordan IK, McDonald JF (1998) Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. *J Mol Evol* 47:14–20

- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc Natl Acad Sci USA* 97:6603–6607
- Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* 9:411–412; author reply 414
- Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* 3:e181
- Kavi HH, Fernandez HR, Xie W, Birchler JA (2005) RNA silencing in *Drosophila*. *FEBS Lett* 579:5940–9
- Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H (2008) *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453:793–797
- Kazazian HH Jr (1998) Mobile elements and disease. *Curr Opin Genet Dev* 8:343–350
- Kidwell MG, Kidwell JF, Sved JA (1977) Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits, including mutation, sterility and male recombination. *Genetics* 86:813–833
- Kidwell MG (1977) Reciprocal differences in female recombination associated with hybrid dysgenesis in *Drosophila melanogaster*. *Genet Res* 30:77–88
- Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827–1830
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B (1988) On the role of unequal exchange in the containment of transposable element copy number. *Genet Res* 52:223–235
- Lankenau S, Corces VG, Lankenau DH (1994) The *Drosophila* micropia retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. *Mol Cell Biol* 14:1764–1775
- Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W (1990) DNA sequence comparison of micropia transposable elements from *Drosophila hydei* and *Drosophila melanogaster*. *Chromosoma* 99:111–117
- Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W (1988) Micropia : a retrotransposon of *Drosophila* combining structural features of DNA viruses, retroviruses and non-viral transposable elements. *J Mol Biol* 204:233–246
- Le Rouzic A, Decelière G (2005) Models of the population genetics of transposable elements. *Genet Res* 85:171–181
- Lerat E, Capy P, Biémont C (2002a) The relative abundance of dinucleotides in transposable elements in five species. *Mol Biol Evol* 19:964–967
- Lerat E, Capy P, Biémont C (2002b) Codon usage by transposable elements and their host genes in five species. *J Mol Evol* 54:625–637
- Lerat E, Rizzon C, Biémont C (2003) Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. *Genome Res* 13:1889–1896
- Lim AK, Kai T (2007) Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 104:6714–6719
- Lipatov M, Lenkov K, Petrov DA, Bergman CM (2005) Paucity of chimeric gene-transposable element transcripts in the *Drosophila melanogaster* genome. *BMC Biol* 3:24
- Lyko F (2001) DNA methylation learns to fly. *Trends Genet* 17:169–72
- Lyko F, Ramsahoye BH, Jaenisch R (2000) DNA methylation in *Drosophila melanogaster*. *Nature* 408:538–40

- Macas J, Neumann P, Navratilova A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8:427
- Macpherson JM, González J, Witten DM, Davis JC, Rosenberg NA, Hirsh AE, Petrov DA (2008) Nonadaptive explanations for signatures of partial selective sweeps in *Drosophila*. Mol Biol Evol 25:1025–1042
- Malik HS, Eickbush TH (2001) Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res 11:1187–1197
- Malik HS, Burke WD, Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805
- Malik HS, Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J Virol 73:5186–5190
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembien LA, Berka J, Braverman, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380
- Marhold J, Rothe N, Pauli A, Mund C, Kuehle K, Brueckner B, Lyko F (2004) Conservation of DNA methylation in dipteran insects. Insect Mol Biol 13:117–123
- Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 24:800–812
- Mathews LM, Chi SY, Greenberg N, Ovchinnikov I, Swergold GD (2003) Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. Am J Hum Genet 72:739–748
- Matranga C, Zamore PD (2007) Small silencing RNAs. Curr Biol 17:R789–93
- Matyunina LV, Jordan IK, McDonald JF (1996) Naturally occurring variation in copia expression is due to both element (cis) and host (trans) regulatory variation. Proc Natl Acad Sci USA 93:7097–102
- McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801
- McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344–355
- McCollum AM, Ganko EW, Barass PA, Rodriguez JM, McDonald JF (2002) Evidence for the adaptive significance of an LTR retrotransposon sequence in a *Drosophila* heterochromatic gene. BMC Evol Biol 2:5
- McDonald JF, Matyunina LV, Wilson S, Jordan IK, Bowen NJ and Miller WJ (1997) LTR retrotransposons and the evolution of eukaryotic enhancers. Genetica 100:3–13
- Mhanni AA, McGowan RA (2004) Global changes in genomic methylation levels during early development of the zebrafish embryo. Dev Genes Evol 214:412–417
- Miller WJ, McDonald JF, Pinsker W (1997) Molecular domestication of mobile elements. Genetica 100:261–270
- Morgan TH, Sturtevant AH, Muller HJ, Bridges CB (1915) The mechanisms of Mendelian heredity. H Holt and Co, New York
- Mugnier N, Biémont C, Vieira C (2005) New regulatory regions of *Drosophila* 412 retrotransposable element generated by recombination. Mol Biol Evol 22:747–757
- Mugnier N, Gueguen L, Vieira C, Biémont C (2008) The heterochromatic copies of the LTR retrotransposons as a record of the genomic events that have shaped the *Drosophila melanogaster* genome. Gene 411:87–93
- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z et al (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316:1718–1723

- Nuzhdin SV, Mackay TFC (1995) The genomic rate of transposable elements movement in *Drosophila melanogaster*. *Mol Biol Evol* 12:180–181
- Pane A, Wehr K, Schüpbach T (2007) *Zucchini* and *squash* encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell* 12:851–862
- Pardue ML, Gall JG (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc Natl Acad Sci USA* 64:600–604
- Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7:597–606
- Pélisson A, Sarot E, Payen-Groschene G, Bucheton A (2007) A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J Virol* 81:1951–60
- Picard G (1976) Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of the I factor. *Genetics* 1983:107–123
- Ponger L, Li WH (2005) Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 22:1119–1128
- Pradhan S, Esteve PO (2003) Mammalian DNA (cytosine-5) methyltransferases and their expression. *Clin Immunol* 109:6–16
- Rebollo R, Lerat E, Kleine Lopez L, Biémont C, Vieira C (2008) Losing *helenia*: the extinction of a *Drosophila* LINE-like element. *BMC Genomics* 9:149
- Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447:425–32
- Richards EJ (2008) Population epigenetics. *Curr Opin Genet Dev* 18:221–226
- Saeder H, Starlinger P (1992) Twenty-five years of transposable element research in Köln. In: Fedoroff N, Botstein D (eds) *The dynamic genome*. Cold Spring Harbor, Cold Spring Harbor, pp 243–263
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20:43–45
- Schaefer M, Lyko F (2007) DNA methylation with a sting: an active DNA methylation system in the honeybee. *Bioessays* 29:208–211
- Shapiro JA (1969) Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *J Mol Biol* 40:93–105
- Silva JC, Kidwell MG (2004) Evolution of P elements in natural populations of *Drosophila willistoni* and *D. sturtevantii*. *Genetics* 168:1323–1335
- Simmons MJ, Raymond JD, Grimes CD, Belinco C, Haake BC, Jordan M, Lund C, Ojala TA, Papermaster D (1996) Repression of hybrid dysgenesis in *Drosophila melanogaster* by heat-shock-inducible sense and antisense P-element constructs. *Genetics* 144:1529–1544
- Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8:272–285
- Soper SFC, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P, Bortvin A (2008) Mouse Maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* 15:285–297
- Suh DS, Choi EH, Yamazaki T, Harada K (1995) Studies on the transposition rates of mobile genetic elements in a natural population of *Drosophila melanogaster*. *Mol Biol Evol* 12:748–758
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genetics* 9:465–476
- Taft JR, Mattick JS (2003) Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding sequences. *Genome Biol* 5:P1

- Thompson JN Jr, Woodruff RC (1980) Increased mutation in crosses between geographically separated strains of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 77:1059–1062
- Tran RK, Zilberman D, de Bustos C, Ditt RF, Henikoff JG, Lindroth AM, Delrow J, Boyle T, Kwong S, Bryson TD, Jacobsen SE, Henikoff S (2005) Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. *Genome Biol* 6:R90
- Tribolium Genome Sequencing Consortium (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452:949–955
- Vagin VV, Klenov MS, Kalmykova AI, Stolyarenko AD, Kotelnikov RN and Gvozdev VA (2004) The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol* 1:54–8
- Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313:320–4
- Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW and Martienssen RA (2007) Epigenetic Natural Variation in *Arabidopsis thaliana*. *PLoS Biol* 5:e174
- Vieira C, Biémont C (2004) Transposable element dynamics in two sibling species *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* 120:115–123
- Vieira C, Piganeau G, Biémont C (2000) High copy numbers of multiple transposable element families in an Australian population of *Drosophila simulans*. *Genet Res* 76:117–119
- Vieira C, Lepetit D, Dumont S, Biémont C (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol* 16:1251–1255
- Vieira C, Aubry P, Lepetit D, Biémont C (1998) A temperature cline in copy number for 412 but not roo/B104 retrotransposons in populations of *Drosophila simulans*. *Proc Biol Sci* 265:1161–1165
- Vieira C, Biémont C (1997) Transposition rate of the 412 retrotransposable element is independent of copy number in natural populations of *Drosophila simulans*. *Mol Biol Evol* 14:185–188
- Vieira C, Biémont C (1996) Geographical variation in insertion site number of retrotransposon 412 in *Drosophila simulans*. *J Mol Evol* 42:443–451
- Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* 28:913–922
- Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE (2006) Functional CpG methylation system in a social insect. *Science* 314:645–647
- Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. *Curr Opin Genet Dev* 18:188–192
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973–982
- Wilson S, Matyunina LV, McDonald JF (1998) An enhancer region within the copia untranslated leader contains binding sites for *Drosophila* regulatory proteins. *Gene* 209:239–46
- Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, et al (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* 306:1937–1940
- Zhai J, Liu J, Liu B, Li P, Meyers BC, Chen X, Cao X (2008) Small RNA-directed epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Genet* 4:e1000056

Transposons and the Dynamic Genome

Lankenau, D.; Volff, J.-N. (Eds.)

2009, XV, 184 p., Hardcover

ISBN: 978-3-642-02004-9